SHORTER ARTICLES AND DISCUSSION

THE APPLICATION OF CORRELATION FORMULÆ TO THE PROBLEM OF VARIETAL DIFFERENCES IN DISEASE RESISTANCE: DATA FROM THE VERMONT EXPERIMENTS WITH POTATOES

THE ultimate practical object of any study of disease resistance in a series of varieties is the selection for future cultivation of the few which are least susceptible. In practise a relatively large series of varieties or strains is taken into cultivation for preliminary study. The size of the cultures of the individual strains must, on a given area, be inversely proportional to their number. Since the individual cultures are necessarily small, it is impossible to assert from the results of a single test that the observed differences between the strains really represent varietal differences in disease resistance. They may be due merely to inadequately large cultures or to imperfectly controlled experimental conditions. It is therefore necessary to repeat the experiment another year or in a different locality in order to determine whether the observed differences are really persistent, and so characteristic of the strain, or whether they are due to transient conditions only. The problem is then purely and simply one of correlation. This is obviously true whether one chooses to avail himself of the advantages of the statistical formulæ or not. If the correlation between disease incidence in cultures of the series of varieties grown in different years, or places, be zero, the varieties show no permanent differentiation in disease resistance. If the correlation has a significant positive value it indicates at once that there are really inherent varietal differences in disease resistance. The numerical magnitude of the correlation indicates something of the extent of this differentiation. If the correlation be low, the prospect of isolating varieties sensibly more resistant that the average will be slight. If the correlation be high, it should be relatively easy to secure highly resistant strains.

Since the correlation method seems to have considerable value in the analysis of data of this kind, I have thought it might be of service to geneticists and plant pathologists to illustrate it by the constants which I have found it necessary to deduce for another purpose from the published records of the series of experiments on disease resistance in varieties of potatoes carried on during

[1] Stuart, W., "Disease Resistance in Potatoes." *Bull. Vt. Agr. Exp. Sta.*, 179, 1914.

the past several years at the Vermont Agricultural Experiment Station.

In a recent bulletin Stuart[1] summarizes the data obtained during five years' observations on percentage infection by early blight (*Alternaria solani*). In his Table I he gives the estimated percentage infection in a series of varieties during the period. Since during a portion of the experiment all the varieties were not considered, I have calculated the correlations in two groups. In one case $N = 149$, in the other $N = 50$. The smaller group comprises only varieties also included in the larger. The correlations, calculated by the usual product moment method[2] without grouping, appear in the accompanying table.[3]

CORRELATION FOR VINE RESISTANCE TO EARLY BLIGHT IN TWO YEARS

Years Compared	Series of 149 Varieties	Series of 50 Varieties
1905–1906	$+.055 \pm .055$	$-.056 \pm .095$
1905–1907	$+.438 \pm .045$	$+.420 \pm .079$
1905–1908	———	$-.021 \pm .095$
1906–1907	$+.042 \pm .055$	$+.226 \pm .091$
1906–1908	———	$+.323 \pm .085$
1907–1908	———	$+.082 \pm .095$

Only 2 of the 9 constants are negative; these are insignificant in comparison with their probable errors. All the constants which have substantial values and are materially larger than their probable errors are positive in sign. The average of the two negative constants is $-.038$, of the seven positive coefficients $+.227$, and of all the (unweighted) values $+.168$. Thus there is clearly a measurable differentiation of the varieties in respect to susceptibility to *Alternaria*.

The values are, however, exceedingly variable, ranging as they do from $-.056$ to $+.438$. The great variation in the actual constants I am inclined to attribute to (*a*) the difficulty of estimating the percentage of infection, (*b*) the unavoidable experimental errors associated with relatively small cultures, and (*c*) the wide variation in average percentage infection from year to year. Both (*a*) and (*b*) are factors which tend to render the actually recorded percentages somewhat erroneous as measures of the real susceptibility of the variety, and tend in consequence to dilute the strength of the correlation. With respect to the third

[2] AMER. NAT., 44: 693–699, 1910.

[3] The chief discrepancy between the results for the larger ($N = 149$) and the smaller ($N = 50$) series of varieties is to be seen in the interrelationship for 1906 and 1907 where the two correlations are $.042 \pm .055$ and $.226 \pm .091$. Here the disagreement is apparent rather than real. The difference is $.184 \pm .106$, which can not be considered statistically trustworthy.

factor, (*c*), it is obvious that if there be only a very slight average percentage infection the test of disease resistance will not be a very critical one, whereas the average percentage can not be very high indeed unless conditions are so unfavorable that all varieties are affected. The percentage infection of early blight varies enormously from year to year. Thus:

Year	Percentage Infection $N = 149$	$N = 50$
1905	3.3	2.4
1906	85.0	83.2
1907	41.7	38.9
1908	—	10.8

With an incidence of 3 per cent. one year and of 85 per cent. the following season one can, in view of the considerations mentioned above, hardly expect to obtain smooth values of the correlation coefficient.

It is interesting to compare these results with those for other maladies of the potato. In the same publication Stuart gives the results of trials for resistance of tubers to scab. Unfortunately the experiments of the second year, 1907, included only 20 of the 65 varieties from the first year. Calculations may be based on the percentage of tubers which are free or nearly free from scab. This is much lower the second year.

	1906	1907
Mean	64.21	28.20
S. D.	11.59	16.33
Correlation	$r = .591 \pm .098$	

The probable error is high because of the fewness of the varieties retained in the second year's test, but the correlation is of more than medium value and is relatively about 6 times as large as its probable error. Thus susceptibility to scab is probably to a very considerable extent a varietal character.

The results for tuber rot tests are not available for successive years, but Stuart has given[4] the percentage of tuber rot in 89 varieties grown on sandy loam and on clay loam soil in 1905. For these I find

	Sandy Loam	Clay Loam
Mean	8.68	39.76
S. D.	11.28	26.85
Correlation	$r = .653 \pm .041$[5]	

[4] Stuart, W., "Disease Resistance in Potatoes," *Bull. Vt. Agr. Exp. Sta.*, 122, Tables VI–VII, 1906.

[5] The value of r given as $.707 \pm .045$ in *Science*, N. S., 38: 402–403, 1913, is deduced from the 62 varieties for which laboratory cultures were available, and from modified percentages. The values agree within the limits of their probable errors.

The correlation of these data, somewhat smoothed by Jones,[6] with a series of determinations of the percentage growth of the fungus on tubers *in the laboratory* has already been determined.[7] For laboratory growth and loss on clay loam, $r = .584 \pm .059$. For laboratory growth and loss on sandy loam, $r = .594 \pm .055$.

Taken as a whole these correlations indicate (*a*) that susceptibility to both early and late blight and to scab differs greatly from variety to variety, and (*b*) that, so far as the evidence goes, the varieties differ more in resistance to tuber injury than to foliage infection by early blight.

It is not at all necessary that the correlations be drawn between the amount of injury to the same organs of the plant or by the same disease. In many instances the so-called cross correlations yield valuable results.

For example Stuart[8] discusses the question of the relationship between vine infection and tuber rot. The point may be subjected to a statistical test by correlating between the maximum percentage of foliage affected by late blight as given in his Table V for potatoes grown on sandy loam soil in 1905 and percentage of rot as recorded in his Tables VI and VII. Unfortunately, the percentages are available for the vines for sandy loam soil only (Table V) while the figures for tuber rot are given for both sandy loam and clay loam soil. Both correlations may be worked out. I find:

For percentage foliage infection on sandy loam soil and per cent. tuber rot on sandy loam soil

$$N = 131, \quad r = .316 \pm .053.$$

For percentage foliage infection on sandy loam soil and per cent. tuber rot on clay loam soil

$$N = 80, \quad r = .102 \pm .075.$$

In both cases the correlations are positive, and hence such evidence as they furnish indicates that the varieties which show the greatest infection of the leaves actually are the worst to rot. That the correlation between injury to the tops and tuber rot is higher on the sandy loam soil is not at all surprising, since the same individual plants—not merely the same varieties—are in-

[6] Jones, L. R., N. J. Giddings and B. F. Lutman, "Investigations of the Potato Fungus *Phytophthora infestans*," *Bull. Vt. Agr. Exp. Sta.*, 168: 74–81, 1912.

[7] Jones and collaborators, *loc. cit.*, and the Reviewer, *Science*, N. S., 38: 402–413, 1913.

[8] *Bull. Vermont Agr. Exp. Sta.*, 122, p. 116.

volved in the correlation. The problem is, however, a complicated one and much more extensive data are needed for a complete analysis.

A problem of very great biological interest as well as of practical importance is that of the specificity of disease resistance. Concretely: Do varieties differ in their susceptibility to a specific disease only, or do they differ merely in susceptibility to disease in general?

A comprehensive and final answer will require far more data than are available and more stringent statistical analysis than can be illustrated here. Some progress can be made by the method of correlation as follows.

If susceptibility be purely *specific* there should be no correlation between the incidence of disease x in year (or culture) p and disease y in year (or culture) q, although there should be a correlation between the incidence of disease x or disease y in different years or cultures. If, on the other hand, differences in disease resistances from variety to variety are determined solely by general weakness or vigor of the stocks, one should expect the correlations between the incidence of different diseases in different years or cultures to be (within the limits fixed by the errors of measurement and the probable errors of random sampling) as high as those between two series of determinations of incidence of one and the same parasite.

Consider first the relationship between the percentage of foliage injury by early blight in 1905, 1906, and 1907 and the percentage of tuber rot in 1905. The correlations are:

Foliage Injury	Per Cent. Tuber Rot on Sandy Loam, 1905 $N = 131$	Per Cent. Tuber Rot on Clay Loam, 1905 $N = 89$
Early Blight, 1905....	$.167 \pm .057$	$.256 \pm .057$
Early Blight, 1906....	$.211 \pm .056$	$.249 \pm .067$
Early Blight, 1907....	$.291 \pm .054$	$.440 \pm .058$

Without exception the correlations are positive in sign. While numerically low, the most of them taken individually may be considered statistically significant in comparison with their probable errors.

Thus it seems clear that the varieties with foliage most injured by early blight are also most subject to tuber rot, just as has been shown to be the case in foliage and tuber infection by late blight.

For foliage infection by early blight in 1905, 1906, and 1907 and foliage injury by late blight in 1905 I find:

	Correlation, $N = 131$
Early Blight, 1905, and Late Blight, 1905..	$-.066 \pm .059$
Early Blight, 1906, and Late Blight, 1905..	$+.190 \pm .057$
Early Blight, 1907, and Late Blight, 1905..	$-.040 \pm .059$

The results are not so consistent as those of the preceding table. The two negative coefficients are insignificant in comparison with their probable errors, and the positive one is not large, either absolutely or relatively. Possibly the laxness of the correlation is in part due to the fact that the measurement of both characters is subject to a large possible error.[9]

For freedom of the tubers from scab with the incidence of other diseases every possible correlation has been determined. The coefficients are shown in the accompanying table. Note that in this case the correlation is between *freedom* from one disease and *occurrence* of another disease. Hence a negative coefficient has the same meaning as a positive one in the foregoing discussions.

FOR PERCENTAGE OF TUBERS FREE OR NEARLY FREE FROM SCAB IN 1906 AND 1907 AND INCIDENCE OF OTHER DISEASES

Diseases Compared with Freedom from Scab	Correlation 1906	Correlation 1970
Per Cent. rot on Sandy Loam Soil, 1905	$-.136 \pm .077$ ($N = 74$)	$-.357 \pm .132$ ($N = 20$)
Per Cent. rot on Clay Loam Soil, 1905	$-.280 \pm .096$ ($N = 42$)	$-.030 \pm .187$ ($N = 13$)
Top Injury by Late Blight, 1905	$+.035 \pm .084$ ($N = 65$)	$-.224 \pm .151$ ($N = 18$)
Top Injury by Early Blight, 1905	$-.561 \pm .054$ ($N = 74$)	$+.240 \pm .142$ ($N = 20$)
Top Injury by Early Blight, 1906	$-.119 \pm .077$ ($N = 74$)	$-.352 \pm .132$ ($N = 20$)
Top Injury by Early Blight, 1907	$-.118 \pm .077$ ($N = 74$)	$-.040 \pm .151$ ($N = 20$)

Because of the small number of varieties involved and the roughness of the measurements the correlations are low and irregular. In ten cases the negative sign indicates that the varieties which are most free from scab are also least susceptible to attacks by other diseases. In neither of the two cases of positive correlation is the constant statistically significant in comparison with its probable error.

Thus altogether 23 of these cross correlations—that is correlations between injury to different organs by the same disease, or

[9] In one case, 1905, the correlation is between foliage injury by two different organisms *in the same year*. What interrelationship is to be expected in such case requires further consideration.

to the same organ by different diseases, or to different organs by different diseases—have been worked out. Only 4 of these—that is only about one case out of six—are exceptions to the rule that varieties which show more than the average amount of injury by one disease will, on the whole, show more than the average injury by another disease. No one of these exceptional constants can be considered significant with regard to its probable error. Several of the 19 which indicate the rule may be looked upon as individually trustworthy. Thus notwithstanding the large variations in numerical magnitude incident to small series of data and rough measurement, the determinations taken collectively certainly furnish highly convincing evidence that to a considerable extent susceptibility to disease is general rather than specific.

The fact that the series of correlation coefficients here presented justify much more definite conclusions than those who have considered the data without statistical analysis have drawn, is sufficient indication of the usefulness of the biometric method in the preliminary stage of disease-resistance experiments in which large numbers of strains are being tested, and in which the mass of data is highly confusing. The special cases illustrated by no means exhaust the possibilities of the biometric formulæ now available. Had the data been more extensive, the analysis might have been carried much farther.

Nothing that has been said in this paper in emphasis of the statistical method must be taken to imply that the most careful individual analysis is not desirable and essential. The two methods are not mutually exclusive, but supplemental.

J. ARTHUR HARRIS

THE DIFFERENT MEANINGS OF THE TERM "FACTOR" AS AFFECTING CLEARNESS IN GENETIC DISCUSSION[1]

IN the analysis of alternative (or segregating) heredity, we find that certain potentialities, such as that of producing a certain color in some part of the soma, appear to be inherited independently of certain other potentialities. We assume that the germ-plasm carries various corresponding genes, factors, or determiners, whose independence in gametogenesis determines the independence of the somatic characters. Cytological study leads

[1] Paper No. 39, University of California, Citrus Experiment Station, Riverside, California.